

# Volterra filters for quantum estimation and detection

Mankei Tsang<sup>1,2,\*</sup>

<sup>1</sup>*Department of Electrical and Computer Engineering,  
National University of Singapore, 4 Engineering Drive 3, Singapore 117583*

<sup>2</sup>*Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551*  
(Dated: December 15, 2015)

The implementation of optimal statistical inference protocols for high-dimensional quantum systems is often computationally expensive. To avoid the difficulties associated with optimal techniques, here I propose an alternative approach to quantum estimation and detection based on Volterra filters. Volterra filters have a clear hierarchy of computational complexities and performances, depend only on finite-order correlation functions, and are applicable to systems with no simple Markovian model. These features make Volterra filters appealing alternatives to optimal nonlinear protocols for the inference and control of complex quantum systems. Applications of the first-order Volterra filter to continuous-time quantum filtering, the derivation of a Heisenberg-picture uncertainty relation, quantum state tomography, and qubit readout are discussed.

## I. INTRODUCTION

The advance of quantum technologies relies on our ability to measure and control complex quantum systems. An important task in quantum control is to infer unknown variables from the noisy measurements of a quantum system. Examples include the prediction of quantum dynamics for measurement-based feedback control [1–5] and the estimation and detection of weak signals [6–23]. To implement the signal processing for such tasks, a Bayesian decision-theoretic formulation of optimal quantum statistical inference is now well established [1–7, 17–19]. The quantum filtering theory pioneered by Belavkin [24, 25] for the optimal prediction of quantum dynamics has especially been hailed as a seminal achievement in quantum control theory; its applications to measurement-based cooling [26], squeezing [27], state preparation [28], quantum error correction [29, 30], qubit readout [20–22], and quantum state tomography [11–14] in atomic, optical, optomechanical, condensed-matter, and superconducting-microwave-circuit systems [1] have been studied extensively in the literature.

Although optimal quantum inference has been successful experimentally for low-dimensional systems, such as qubits [31] and few-photon systems [32], as well as near-Gaussian systems, such as optical phase estimation [23] and optomechanics [33], its implementation for high-dimensional non-Gaussian quantum systems is beset with difficulties in practice. An exact implementation of the quantum Bayes rule [2] for optimal inference requires numerical updates of the posterior density matrix based on the measurement record. Except for special cases such as Gaussian systems [1], the number of elements needed to keep track of the density matrix scales exponentially with the degrees of freedom, making the implementation prohibitive for many-body non-Gaussian systems. This problem, known as the curse of dimensionality, means

that approximations must often be sought [26, 29, 30, 34–37]. Current approximation techniques for dynamical systems include Gaussian approximations [13, 26, 34], phase-space particle filters [36], Hilbert-space truncation [30, 35], and manifold learning [37], but these techniques provide little assurance about their actual errors and often remain too expensive to compute for real-time control of high-dimensional systems. Another problem with optimal inference and the associated stochastic-master-equation approach is its reliance on a Markovian model, which is difficult to use for many complex systems, especially those with  $1/f$  or fractional noise statistics. With the ongoing trend of increasing complexity in quantum experiments, not only with condensed matter but also with optomechanics [38], atomic ensembles [39], and superconducting circuits [40], optimal inference is becoming an unattainable goal in practice.

Against this backdrop, here I propose an alternative approach to quantum estimation and detection based on Volterra filters. Instead of seeking absolute optimality, Volterra filters are a class of polynomial estimators with a clear hierarchy of computational complexities and estimation errors [41]. Their applications to quantum estimation and detection promise to solve many of the practical problems associated with optimal quantum inference, including the curse of dimensionality, the lack of error assurances upon approximations, and the need for a Markovian model. The filter errors also provide a set of upper error bounds on the Bayesian quantum Cramér-Rao [6, 7, 42], Ziv-Zakai [43], and Helstrom [6, 7, 44] bounds, forming novel hierarchies of fundamental uncertainty relations and may be of independent foundational interest. The Volterra series has recently been used to model the input-output relations of a quantum system [45], but my focus here is different and concerns the estimation of hidden observables and hypothesis testing given the output measurement record.

\* mankei@nus.edu.sg

## II. QUANTUM ESTIMATION

### A. Formalism

Consider a quantum system in the Heisenberg picture with initial density operator  $\rho$ . Let

$$y = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(K) \end{pmatrix} \quad (2.1)$$

be a column vector of observables under measurement. For example,  $y$  can be the observables of an output optical field under homodyne, heterodyne, or photon-counting measurements. Given a measurement record of  $y$ , the goal of quantum estimation is to infer a column vector of hidden observables

$$x \equiv \begin{pmatrix} x(1) \\ x(2) \\ \vdots \\ x(J) \end{pmatrix}. \quad (2.2)$$

For example,  $x$  can be the observables of a quantum system that has interacted with the optical field, such as the position of a quantum mechanical oscillator or a spin operator of an atomic ensemble, and the goal of the estimation is to infer  $x$  given the measurement record. Quantum estimation is usually framed in the Schrödinger picture via the concept of posterior density operator [1, 2], but it can be shown to be equivalent to the Heisenberg-picture approach adopted here [4]. This task is especially important for measurement-based feedback control [1], such as measurement-based cooling and squeezing, to gain real-time information about quantum degrees of freedom and to reduce their uncertainties via feedback control. Experiments that implement quantum estimation have been reported in Refs. [31–33] for example.

The estimation error has a well-defined decision-theoretic meaning if all the  $x$  and  $y$  operators commute with one another, such that  $x$  and  $y$  can be jointly measured and treated as classical random variables in the same probability space [4, 7, 46]. This assumption is applicable to a wide range of scenarios, including quantum filtering [4, 46] and the estimation of any classical parameter or waveform coupled to a quantum system [17, 47].

Since  $x$  and  $y$  are compatible observables, the rest of the estimation theory is identical to the classical treatment [41]. Let  $\tilde{x}(j|y)$  be an estimator of  $x(j)$  given  $y$ , and assume that the estimator is given by the truncated Volterra series, viz.,

$$\begin{aligned} \tilde{x}(j|y) = & \sum_{p=0}^P \sum_{1 \leq k_1 \leq k_2 \leq \dots \leq k_p \leq K} h_p(j, k_1, k_2, \dots, k_p | \theta) \\ & \times y(k_1) y(k_2) \dots y(k_p), \end{aligned} \quad (2.3)$$

where  $\theta$  is a vector of tunable parameters,  $P$  is the order of the series and quantifies the complexity of the filter, and the zeroth-order term is simply a constant  $h_0(j)$  and does not depend on  $y$ . For  $P \rightarrow \infty$ , the series can be regarded as the Taylor series for an arbitrary estimator, although I will focus on finite  $P$ .

A useful trick to simplify the notations is to define the set of all products of  $y$  elements up to order  $P$  as

$$y^{(P)} \equiv \{1, y, y^{\otimes 2}, \dots, y^{\otimes P}\}, \quad (2.4)$$

where

$$\begin{aligned} y^{\otimes p} \equiv & \{y(k_1)y(k_2)\dots y(k_p); \\ & 1 \leq k_1 \leq k_2 \leq \dots \leq k_p \leq K\} \end{aligned} \quad (2.5)$$

is the set of all  $p$ th-order products of  $y$  elements. Then the Volterra series in Eq. (2.3) can be rewritten as

$$\tilde{x}(j|y) = \sum_{\mu} h^{(P)}(\mu | \theta) y^{(P)}(\mu), \quad (2.6)$$

where  $h^{(P)}$  is a linear filter with respect to  $y^{(P)}$  but equivalent to the Volterra filter that is nonlinear with respect to  $y$ , and  $\mu$  is a composite index that goes through all elements in  $y^{(P)}$ .

Define

$$\langle f(x, y) \rangle \equiv \text{tr} [\rho f(x, y)] \quad (2.7)$$

as the expectation of any function of  $x$  and  $y$ , with  $\text{tr}$  denoting the operator trace. Let the error covariance matrix be

$$\Sigma(j, k) \equiv \langle [x(j) - \tilde{x}(j|y)] [x(k) - \tilde{x}(k|y)] \rangle. \quad (2.8)$$

The absolutely minimum mean-square error for arbitrary estimators is achieved by the conditional expectation of  $x$  given  $y$  [4]. For the optimal filtering and prediction of quantum observables for example, the usual method is to compute the posterior density operator  $\rho(y)$  conditioned on the measurement record  $y$  in the Schrödinger picture using the Kraus operators that characterize the measurements [1, 2], and then take the conditional expectation given by  $\tilde{x}(j|y) = \text{tr}[x_S(j)\rho(y)]$ , with  $x_S(j)$  being the Schrödinger picture of  $x(j)$ . If the continuous-time limit is taken, the posterior density operator obeys the celebrated stochastic master equation [1–4] first proposed by Belavkin [24, 25]. The computation of  $\rho(y)$  suffers from the curse of dimensionality however. To restrict the complexity, consider here instead the error of the  $P$ th-order

Volterra filter given by

$$\begin{aligned}
\Sigma^{(P)}(j, k|\theta) &= \left\langle \left[ x(j) - \sum_{\mu} h^{(P)}(j, \mu|\theta) y^{(P)}(\mu) \right] \right. \\
&\quad \times \left. \left[ x(k) - \sum_{\mu} h^{(P)}(k, \mu|\theta) y^{(P)}(\mu) \right] \right\rangle \\
&= C_x(j, k) - \sum_{\mu} h^{(P)}(j, \mu|\theta) C_{xy^{(P)}}(k, \mu) \\
&\quad - \sum_{\mu} h^{(P)}(k, \mu|\theta) C_{xy^{(P)}}(j, \mu) \\
&\quad + \sum_{\mu, \nu} h^{(P)}(j, \mu|\theta) h^{(P)}(k, \nu|\theta) C_{y^{(P)}}(\mu, \nu),
\end{aligned} \tag{2.9}$$

where

$$C_x(j, k) \equiv \langle x(j)x(k) \rangle, \tag{2.11}$$

$$C_{xy^{(P)}}(j, \mu) \equiv \langle x(j)y^{(P)}(\mu) \rangle, \tag{2.12}$$

$$C_{y^{(P)}}(\mu, \nu) \equiv \langle y^{(P)}(\mu)y^{(P)}(\nu) \rangle. \tag{2.13}$$

To optimize the Volterra filter, one can seek the parameters  $\theta$  that minimize any desired component of  $\Sigma^{(P)}(j, k|\theta)$  in Eq. (2.10), which has the remarkable feature of depending only on finite-order correlations. Specifically,  $C_{xy^{(P)}}(j, \mu)$  depends on the correlation between  $x(j)$  and products of  $y$  elements up to the  $P$ th order, and  $C_{y^{(P)}}$  depends on the correlations among  $y$  up to the  $2P$ th order. Stationarity assumptions and frequency-domain techniques can further simplify the expressions.

Quantum mechanics comes into the problem through the correlations. They must obey uncertainty relations with other incompatible observables [7, 48]. They can violate Bell [49] and Leggett-Garg [50] inequalities, requiring different probability spaces for different experimental settings. They may result from nontrivial internal quantum dynamics with no classical correspondence; the promise of quantum computation and simulation [51] is in fact based on the difficulty of reproducing quantum dynamical statistics using any hidden-variable model. This difficulty also means that attempts to simplify quantum filters via classical models [26, 34, 36] are likely to be inaccurate for highly nonclassical systems. The Volterra filters sidestep the issue via a manifestly non-Markovian approach that does not require an *online* simulation of the internal quantum dynamics. The identification of the correlations and the filter synthesis, though nontrivial, can be done *offline* for control applications.

A challenge for classical applications of Volterra filters is that the correlations are often difficult to model or measure in practice, but it is less problematic for quantum systems: computing and measuring correlation functions is already a major endeavor in condensed-matter physics [52] and early quantum optics [53] with an extensive literature. The Volterra-series approach to input-output

analysis [45] should also help their simulation. Compared with the stochastic-master-equation approach [1–4], the use of correlation functions has the advantage of not requiring a Markovian model or stochastic calculus, although the Volterra filters may require a longer memory depending on the time scales of the correlation functions and the signal-to-noise properties. An empirical alternative to prior system identification is to train the filter directly using experimental or simulated data to minimize the sample errors.

I now consider the ideal case where arbitrary Volterra filters can be implemented, such that the tunable parameters  $\theta$  are all elements of  $h^{(P)}$ . Since  $\Sigma^{(P)}$  is quadratic with respect to  $h^{(P)}$ , the minimization can be performed analytically. Define the risk function [54] to be minimized as

$$R(\theta) \equiv \sum_{j,k} u(j) \Sigma^{(P)}(j, k|\theta) u(k), \tag{2.14}$$

where  $u$  is an arbitrary real vector. The optimal Volterra filter

$$\tilde{h}^{(P)} \equiv \arg \min_{h^{(P)}} R(h^{(P)}) \tag{2.15}$$

for arbitrary  $u$  satisfies the equation

$$C_{xy^{(P)}}(j, \nu) = \sum_{\mu} \tilde{h}^{(P)}(j, \mu) C_{y^{(P)}}(\mu, \nu), \tag{2.16}$$

which is a system of linear equations with respect to  $\tilde{h}^{(P)}$  and can be solved by conventional methods, and the resulting error covariance matrix is

$$\tilde{\Sigma}^{(P)}(j, k) \equiv \Sigma^{(P)}(j, k|\tilde{h}^{(P)}) \tag{2.17}$$

$$= C_x(j, k) - \sum_{\mu} \tilde{h}^{(P)}(j, \mu) C_{xy^{(P)}}(k, \mu). \tag{2.18}$$

This error can be computed offline to evaluate the optimal performance of a Volterra filter and the trade-off between the error and the filter complexity  $P$ . Going to a higher order is guaranteed not to increase the error, since  $\tilde{\Sigma}^{(P)} \leq \tilde{\Sigma}^{(Q)}$  if  $P > Q$  (a higher-order filter can always achieve the performance of a lower-order filter by ignoring the higher-order terms in  $y^{(P)}$ ). As the infinite-order Volterra filter can be regarded as the Taylor series for an arbitrary function,  $\tilde{h}^{(\infty)}$  will be the optimal among arbitrary estimators and  $\tilde{\Sigma}^{(\infty)}$  will coincide with the absolutely optimal error.  $\tilde{\Sigma}^{(P)}$  thus provides a hierarchy of increasingly tight upper error bounds for optimal quantum inference. Most importantly, a finite-order Volterra filter can still enjoy a performance given by Eq. (2.18) for any statistics, even if it is not optimal in the absolute sense. On a fundamental level, it is interesting to note that, if  $x$  is classical, the upper error bounds also apply to the Bayesian quantum Cramér-Rao [6, 7, 42] and Ziv-Zakai [43] lower error bounds, forming a novel

set of operationally motivated uncertainty relations; an example is shown in Sec. II C.

The optimal  $P = 0$  Volterra filter does not process the measurement and is simply given by the prior expectation  $\langle x \rangle$ . The  $P = 1$  Volterra filter is a linear filter with respect to  $y$  and deserves special attention, as it is the simplest Volterra filter beyond the trivial zeroth-order case and will likely become the most popular. If  $x$  and  $y$  are jointly Gaussian, the optimal linear filter is also the optimal among arbitrary estimators and equivalent to the Kalman filter when applied to the prediction of Markovian dynamical systems [55], but the linear filter can still be used for any non-Gaussian or non-Markovian statistics and depends only on the second-order correlations in terms of  $x$  and  $y$ .

### B. Continuous-time quantum filtering

For example, consider the continuous-time quantum filtering and prediction problem, which is to estimate a Heisenberg-picture observable  $x(t)$  given the past measurement record  $\{y(\tau); t_0 \leq \tau \leq T < t\}$  [4]. It can be shown that all the Heisenberg-picture operators under consideration commute with one another under rather general conditions for filtering and prediction [4, 46]. If  $t < T$  is desired for smoothing [17], care should be taken in the modeling to ensure that  $x(t)$  still commutes with  $y$  and an operational meaning of the estimation error exists. For example, a c-number signal, such as a classical force, commutes with all operators by definition.

To transition from the discrete formalism to continuous time, define a discrete time given by

$$t_j = t_0 + j\delta t, \quad (2.19)$$

with initial time  $t_0$ , integer  $j$ , and time interval  $\delta t$ . For infinitesimal  $\delta t$ , the linear  $P = 1$  estimator in the continuous-time limit becomes

$$\tilde{x}(t|y) = h_0(t) + \int_{t_0}^T d\tau h_1(t, \tau) y(\tau), \quad (2.20)$$

where  $\tilde{x}(t|y)$ ,  $h_0(t)$ ,  $h_1(t, \tau)$ , and  $y(\tau)$  are continuous-time versions of  $\tilde{x}(j|y)$ ,  $h_0(j)$ ,  $h_1(j, k)/\delta t$ , and  $y(k)$ , respectively. Eq. (2.20) is a continuous-time limit of the Volterra series in Eq. (2.3) for  $P = 1$ . Assuming zero-mean  $x$  and  $y$  for simplicity and using Eqs. (2.16) and (2.18), the optimal linear filter  $\tilde{h}_1(t, \tau)$  and the corresponding mean-square error  $\tilde{\Sigma}^{(1)}(t, t)$  can be expressed as

$$C_{xy}(t, \tau) = \int_{t_0}^T ds \tilde{h}_1(t, s) C_y(s, \tau), \quad (2.21)$$

$$\tilde{\Sigma}^{(1)}(t, t) = C_x(t, t) - \int_{t_0}^T d\tau \tilde{h}_1(t, \tau) C_{xy}(t, \tau), \quad (2.22)$$

where

$$C_x(t, t) \equiv \langle x^2(t) \rangle, \quad (2.23)$$

$$C_{xy}(t, \tau) \equiv \langle x(t)y(\tau) \rangle, \quad (2.24)$$

$$C_y(t, \tau) \equiv \langle y(t)y(\tau) \rangle \quad (2.25)$$

are the only correlation functions needed to compute both the filter and the error. Although this form of the optimal linear estimator is known in the classical context [55], its applicability to quantum systems with any nonlinear dynamics and non-Gaussian statistics is hitherto unappreciated. Compared with the stochastic master equation, the linear filter can be more easily implemented using fast digital electronics or even analog electronics in practice [23, 56] for measurement-based feedback control, while the implementation of higher-order filters is more involved but can leverage existing digital-signal-processing techniques [41].

### C. Heisenberg-picture uncertainty relation

To demonstrate a side consequence of the Volterra-filter formalism, here I use the analytic error expression for the first-order Volterra filter to derive a quantum uncertainty relation for Heisenberg-picture operators. Consider the Hamiltonian  $\mathfrak{H}(t) = \mathfrak{H}_0(t) - qx(t)$ , where  $q$  is a canonical position operator,  $x(t)$  is a classical force, and  $\mathfrak{H}_0$  is the rest of the Hamiltonian. Suppose that  $\mathfrak{H}_0$  is at most quadratic with respect to canonical position and momentum operators, such that the equations of motion for those operators in the Heisenberg picture are linear. The initial density operator  $\rho$ , on the other hand, can have any non-Gaussian statistics.

Consider an output field quadrature operator  $y(t)$  that commutes with itself at different times in the Heisenberg picture [4]. For example, it can model the homodyne measurement of an output optical field in optomechanics. It can be shown that

$$y(t) = y_0(t) + \int_0^T dt g(t, \tau) x(\tau), \quad (2.26)$$

where

$$g(t, \tau) = \begin{cases} \frac{i}{\hbar} [y_0(t), q_0(\tau)], & t > \tau, \\ 0, & t \leq \tau, \end{cases} \quad (2.27)$$

is the causal c-number commutator and the subscript 0 denotes the interaction picture with respect to the Hamiltonian  $\mathfrak{H}_0$ .

Without loss of generality, assume that  $x(t)$ ,  $y_0(t)$ , and  $q_0(t)$  are zero-mean processes. Consider the estimation of  $x(t)$  using the record  $\{y(\tau); 0 < \tau \leq T\}$ . If  $y_0(t)$  has non-Gaussian statistics, the optimal nonlinear estimator is difficult to derive, but the first-order Volterra filter given by

$$\tilde{x}(t|y) = \int_0^T d\tau h_1(t, \tau) y(\tau) \quad (2.28)$$

can be analyzed more easily. To proceed, it is more convenient to consider discrete time as defined in Eq. (2.19). Regarding  $x$ ,  $y_0$ ,  $y$ , and  $\tilde{x}$  as column vectors and  $g$  and  $h_1$  as matrices, Eqs. (2.26) and (2.28) can be rewritten in matrix form as

$$y = y_0 + \delta t g x, \quad (2.29)$$

$$\tilde{x} = \delta t h_1 y. \quad (2.30)$$

With covariance matrices defined as

$$C_x \equiv \langle x x^\top \rangle, \quad (2.31)$$

$$C_{y0} \equiv \langle y_0 y_0^\top \rangle, \quad (2.32)$$

$$C_{xy} \equiv \langle x y^\top \rangle = \delta t C_x g^\top, \quad (2.33)$$

$$C_y \equiv \langle y y^\top \rangle = \delta t^2 g C_x g^\top + C_{y0}, \quad (2.34)$$

where  $\top$  denotes the matrix transpose, the optimal linear filter becomes

$$\delta t \tilde{h}_1 = C_{xy} C_y^{-1} = \delta t C_x g^\top (\delta t^2 g C_x g^\top + C_{y0})^{-1}, \quad (2.35)$$

and the error covariance matrix becomes

$$\tilde{\Sigma}^{(1)} \equiv \left\langle \left( x - \delta t \tilde{h}_1 y \right) \left( x - \delta t \tilde{h}_1 y \right)^\top \right\rangle \quad (2.36)$$

$$= C_x - \delta t \tilde{h}_1 C_{xy}^\top \quad (2.37)$$

$$= C_x - \delta t^2 C_x g^\top (\delta t^2 g C_x g^\top + C_{y0})^{-1} g C_x \quad (2.38)$$

$$= (C_x^{-1} + \delta t^2 g^\top C_{y0}^{-1} g)^{-1}, \quad (2.39)$$

where the last line uses the matrix inversion lemma [57].

The error covariance can be compared with the Bayesian quantum Cramér-Rao bound derived in Ref. [42]. The quantum bound for Gaussian  $x$  results in a matrix inequality given by

$$\tilde{\Sigma}^{(1)} \geq \left( C_x^{-1} + \frac{4\delta t^2}{\hbar^2} C_{q0} \right)^{-1}, \quad (2.40)$$

where

$$C_{q0}(t_j, t_k) \equiv \frac{1}{2} \langle q_0(t_j) q_0(t_k) + q_0(t_k) q_0(t_j) \rangle. \quad (2.41)$$

Unlike  $x(t)$  and  $y(t)$ ,  $q(t)$  may not self-commute at different times, and the symmetric ordering in the covariance function [58] arises naturally from the derivation of the quantum bound in Ref. [42]. Comparing Eq. (2.39) and Eq. (2.40), it can be seen that the inequality holds only if

$$g^\top C_{y0}^{-1} g \leq \frac{4}{\hbar^2} C_{q0}, \quad (2.42)$$

which is a matrix uncertainty relation between two quantum processes in the Heisenberg picture involving their causal commutator  $g$ . Note that  $y_0$  and  $q$  are canonical phase-space coordinate operators with linear dynamics but need not have Gaussian statistics. The end result

does not involve  $x$  and can be applied to any quantum system that satisfies the stated assumptions beyond the estimation scenario. The estimation procedure nonetheless gives the relation a clear operational meaning.

Eq. (2.42) can be further simplified by assuming linear-time-invariant dynamics and stationary statistics. The result in the continuous long-time limit is a spectral uncertainty relation given by

$$S_{y0}(\omega) S_{q0}(\omega) \geq \frac{\hbar^2}{4} |G(\omega)|^2, \quad (2.43)$$

with the frequency-domain quantities defined by

$$C_{y0}(t, \tau) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{y0}(\omega) \exp[i\omega(t - \tau)], \quad (2.44)$$

$$C_{q0}(t, \tau) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{q0}(\omega) \exp[i\omega(t - \tau)], \quad (2.45)$$

$$g(t, \tau) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} G(\omega) \exp[i\omega(t - \tau)]. \quad (2.46)$$

The spectral relation imposes a lower bound on the noise floor of an output operator  $y_0(t)$  in terms of the spectrum of a noncommuting operator  $q_0(t)$ . For example, the relation can be used to determine the fundamental limit to the noise floor of optical homodyne detection as a function of the mechanical-position power spectral density for a gravitational-wave detector [8, 59]. The inequality can be saturated if the quantum statistics are Gaussian [42].

#### D. Quantum state tomography

For an application in quantum information processing, consider the estimation of parameters in a density matrix, also known as quantum state tomography [9–15]. Assume a  $d \times d$  density matrix of the form

$$\rho_z = \frac{I}{d} + \sum_{\alpha=1}^{d^2-1} z_\alpha E_\alpha, \quad (2.47)$$

where  $I$  is the identity matrix,  $E_\alpha$  is a set of Hermitian, traceless, and orthonormal matrices that satisfy

$$E_\alpha = E_\alpha^\dagger, \quad \text{tr } E_\alpha = 0, \quad \text{tr } E_\alpha E_\beta = \delta_{\alpha\beta}, \quad (2.48)$$

and  $z$  is a column vector of real unknown parameters.  $\rho_z$  is Hermitian and  $\text{tr } \rho_z = 1$  by construction, and the density matrix describes a physical quantum state only if  $\rho_z$  is positive-semidefinite [11]. Measurements can often be modeled as [11]

$$y = Az + y_0, \quad (2.49)$$

where  $y$  is a column vector,  $A$  is a known measurement matrix, and  $y_0$  is a zero-mean noise vector. The main difficulty with the Bayesian estimation protocol [10, 13, 15] is that, owing to the physical-state requirement, the prior



for  $z$  is highly non-Gaussian, while the statistics of  $y_0$  may also be non-Gaussian. With the non-Gaussian statistics and  $d$  scaling exponentially with the degrees of freedom, exact Bayesian estimation of  $z$  would suffer from the curse of dimensionality. Existing approximation techniques include Gaussian approximations [13] and particle filters [15], but their actual estimation errors remain unclear.

The Volterra filters can be used despite the non-Gaussianity of  $z$  or  $y_0$ . Let

$$x = Bz \quad (2.50)$$

be a column vector of parameters to be estimated for a given sampling matrix  $B$ . Note that  $B$  can be a non-square matrix and the number of elements in  $x$  can be much smaller than that in  $z$  if the dimensionality of the latter is a concern. For example, the fidelity between the density matrix and a target pure state [60] can be expressed in this way, in which case  $B$  is a row vector and  $x$  is a scalar. The optimal first-order filter can be expressed as

$$\tilde{x} = B \langle z \rangle + \tilde{h}_1 (y - A \langle z \rangle), \quad (2.51)$$

$$\tilde{h}_1 = BC_z A^\top (AC_z A^\top + C_{y0})^{-1}, \quad (2.52)$$

$$C_z \equiv \langle zz^\top \rangle - \langle z \rangle \langle z \rangle^\top. \quad (2.53)$$

The filter is guaranteed to offer an error covariance matrix given by

$$\tilde{\Sigma}^{(1)} = B (C_z^{-1} + A^\top C_{y0}^{-1} A)^{-1} B^\top. \quad (2.54)$$

The linear complexity and the error guarantee are the main advantages of the Volterra filter. A shortcoming is that, due to noise and the lack of a constraint in the algorithm, the estimate  $\tilde{x}$  may not lead to a positive-semidefinite density matrix. If this is a problem, an obvious remedy is to find the physical  $x$  closest to  $\tilde{x}$  with respect to a distance measure. A more sophisticated way is to compute the posterior distribution over a region near  $\tilde{x}$  with a volume suggested by  $\tilde{\Sigma}^{(1)}$ . If the noise is low enough or the number of trials is large enough such that  $\tilde{\Sigma}^{(1)}$  is small, the region needs to cover a small parameter subspace only, and the curse of dimensionality can be avoided.

The remaining issue is the choice of prior  $\langle z \rangle$  and  $C_z$  in an objective manner. One option is to take one of the commonly used objective priors for  $z$  [11, 15] and compute its moments. For  $d = 2$  and  $z$  being the Bloch vector, the prior moments can be easily calculated by taking advantage of the Bloch spherical symmetry. The computation seems nontrivial for  $d \geq 3$ , but for each  $d$  it needs to be done just once and for all.

The most conservative and arguably paranoid option is to choose a prior that is least favorable to the Volterra filter. Given a prior probability measure  $\pi_z$  on  $z$ , one can define a risk function, such as the Hilbert-Schmidt distance given by

$$R(\pi_z) = \text{tr } \tilde{\Sigma}^{(1)}(\pi_z). \quad (2.55)$$

Then the least favorable prior is one that maximizes the risk while still observing the physical constraint on  $\rho_x$ , that is,

$$\arg \max_{\pi_z; \rho_z \geq 0} R(\pi_z). \quad (2.56)$$

Note that this prior depends in general on the measurement matrix  $A$  as well as the sampling matrix  $B$ . Without the physical constraint, the least favorable  $C_z$  would be infinite, giving

$$\tilde{\Sigma}^{(1)} \leq B (A^\top C_{y0}^{-1} A)^{-1} B^\top, \quad (2.57)$$

and the Volterra filter would become equivalent to the unconstrained maximum-likelihood estimator for Gaussian  $y_0$ . The effect of a finite  $C_z$  is to pull the estimate from the maximum-likelihood value towards the prior  $\langle x \rangle$  via the weighted average given by Eq. (2.51).

### III. QUANTUM DETECTION

#### A. Formalism

Assume two hypotheses denoted by  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . These hypotheses can be about the initial density operator as well as the dynamics and measurements of the quantum system [18]. As before, let the measured Heisenberg-picture observables be  $y$  with commuting elements under both hypotheses. The goal of detection is equivalent to binary hypothesis testing, which is to make a decision on  $\mathcal{H}_0$  or  $\mathcal{H}_1$  based on  $y$ . Applications include force detection [8, 18, 44], fundamental tests of quantum mechanics [18, 19, 38, 61], quantum error correction [1, 29], and qubit readout [20–22]. Prior work on the use of Volterra filters for classical detection focuses on the heuristic deflection criterion [62, 63], but it does not seem to have any decision-theoretic meaning or relationship with the more rigorous criteria of error probabilities [63]. Here I propose a similar performance criterion that is able to provide an upper bound on the average error probability, while still offering a simple design rule for the Volterra filters. To my knowledge the proposed design rule is new also in the context of classical detection theory.

Let  $\lambda(y)$  be a test statistic as a polynomial function of  $y$  similar to Eq. (2.3). For later notational convenience, I will rewrite it as

$$\lambda(y) = h_0 + H^\top Y, \quad (3.1)$$

where the zeroth-order term  $h_0$  is written separately,  $Y$  is a column vector with the elements in  $\{y, y^{\otimes 2}, \dots, y^{\otimes P}\}$  without the constant term 1,  $H$  is a column vector with the corresponding elements in  $h^{(P)}$ , and  $\top$  denotes the transpose. Let  $\langle f(y) \rangle_0$  be the expectation of a function of  $y$  given hypothesis  $\mathcal{H}_0$ , and  $\langle f(y) \rangle_1$  be the expectation given hypothesis  $\mathcal{H}_1$ . Note that the hypotheses can be about the initial density operator, the dynamics, and the definition of  $y$ .

I demand the test statistic to have different expectations for the two hypotheses, viz.,

$$\langle \lambda \rangle_0 \neq \langle \lambda \rangle_1. \quad (3.2)$$

This means that the order  $P$  cannot be arbitrary but must be high enough to result in different expectations. I further demand the expectations to be symmetric around 0, viz.,

$$\langle \lambda \rangle_0 + \langle \lambda \rangle_1 = 0. \quad (3.3)$$

This is accomplished by setting

$$h_0 = -H^\top \bar{Y}, \quad (3.4)$$

$$\bar{Y} \equiv \frac{1}{2} (\langle Y \rangle_0 + \langle Y \rangle_1), \quad (3.5)$$

resulting in

$$\langle \lambda \rangle_1 = -\langle \lambda \rangle_0 = H^\top \Delta, \quad (3.6)$$

$$\Delta \equiv \frac{1}{2} (\langle Y \rangle_1 - \langle Y \rangle_0). \quad (3.7)$$

Without loss of generality, I assume  $\langle \lambda \rangle_1 = H^\top \Delta > 0$ . Consider a threshold test that decides on  $\mathcal{H}_0$  if  $\lambda < 0$  and  $\mathcal{H}_1$  if  $\lambda \geq 0$ . This is commonly expressed as [55]

$$\lambda(y) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 0. \quad (3.8)$$

The average error probability becomes

$$\mathcal{P}_e(H) = \pi_0 \langle 1_{\lambda \geq 0}(y) \rangle_0 + \pi_1 \langle 1_{\lambda < 0}(y) \rangle_1, \quad (3.9)$$

where  $\pi_0$  and  $\pi_1$  are the prior probabilities for the hypotheses and  $1_{\lambda \geq 0}$  and  $1_{\lambda < 0}$  are indicator functions. Since  $\mathcal{P}_e(H)$  in general depends on infinite orders of  $\lambda$  moments, I appeal to the Cantelli inequality [64] to obtain

$$\langle 1_{\lambda \geq 0}(y) \rangle_0 \leq \frac{\langle \lambda^2 \rangle_0 - \langle \lambda \rangle_0^2}{\langle \lambda^2 \rangle_0}, \quad (3.10)$$

and similarly for  $\langle 1_{\lambda < 0}(y) \rangle_1$ . This leads to upper bounds on  $\mathcal{P}_e$  given by

$$\mathcal{P}_e(H) \leq \mathcal{Q}(H) \leq \mathcal{R}(H), \quad (3.11)$$

$$\begin{aligned} \mathcal{Q}(H) &\equiv \frac{\pi_0}{1 + (H^\top \Delta)^2 / (H^\top C_0 H)} \\ &\quad + \frac{\pi_1}{1 + (H^\top \Delta)^2 / (H^\top C_1 H)}, \end{aligned} \quad (3.12)$$

$$\mathcal{R}(H) \equiv \frac{H^\top (\pi_0 C_0 + \pi_1 C_1) H}{(H^\top \Delta)^2}, \quad (3.13)$$

where

$$C_0 \equiv \langle Y Y^\top \rangle_0 - \langle Y \rangle_0 \langle Y \rangle_0^\top, \quad (3.14)$$

$$C_1 \equiv \langle Y Y^\top \rangle_1 - \langle Y \rangle_1 \langle Y \rangle_1^\top \quad (3.15)$$

are the conditional covariance matrices.  $1/\mathcal{R}$  can be regarded an output signal-to-noise ratio and has a similar form to the deflection criterion [62, 63], although  $\mathcal{R}$  has a clearer decision-theoretic meaning as an upper error bound.

The purpose of using  $\mathcal{R}$  rather than  $\mathcal{P}_e$  or  $\mathcal{Q}$  is to define an easy-to-optimize criterion in terms of finite-order correlations. To find the  $\mathcal{R}$ -optimal filter, consider the Cauchy-Schwarz inequality

$$(H^\top \Delta)^2 \leq (H^\top M H) (\Delta^\top M^{-1} \Delta) \quad (3.16)$$

for any positive-definite matrix  $M$ . The inequality is saturated if and only if  $H = \alpha M^{-1} \Delta$  for any constant  $\alpha$ . Setting  $M = \pi_0 C_0 + \pi_1 C_1$ , I obtain

$$\tilde{R} \equiv \min_H \mathcal{R}(H) = \frac{1}{\Delta^\top (\pi_0 C_0 + \pi_1 C_1)^{-1} \Delta}, \quad (3.17)$$

$$\tilde{H} \equiv \arg \min_H \mathcal{R}(H) = \alpha (\pi_0 C_0 + \pi_1 C_1)^{-1} \Delta, \quad (3.18)$$

and the  $\mathcal{R}$ -optimal test statistic  $\tilde{\lambda}(y) \equiv h_0 + \tilde{H}^\top Y$ , taking  $\alpha = 1$  without loss of generality, becomes

$$\tilde{\lambda}(y) = \Delta^\top (\pi_0 C_0 + \pi_1 C_1)^{-1} (Y - \bar{Y}), \quad (3.19)$$

which can then be used in a threshold test. The merits of this approach are similar to those in the estimation scenario: dependence of  $\tilde{\lambda}(y)$  on finite-order correlations  $\Delta$ ,  $C_0$ , and  $C_1$  without relying on a Markovian model, a performance guaranteed by upper bounds  $\mathcal{P}_e(\tilde{H}) \leq \mathcal{Q}(\tilde{H}) \leq \tilde{\mathcal{R}}$  (the actual  $\mathcal{P}_e$  may be much lower), and a hierarchy of decreasing  $\tilde{R}$  versus increasing complexity. For the study of fundamental quantum metrology,  $\mathcal{P}_e(\tilde{H})$ ,  $\mathcal{Q}(\tilde{H})$ , and  $\tilde{\mathcal{R}}$  also provide a set of upper bounds on the Helstrom bound [6, 7, 44].

It is not difficult to show that, if the hypotheses are about the mean of a Gaussian  $y$  and  $C_0 = C_1$ ,  $\tilde{\lambda}(y)$  for  $P = 1$  coincides with the well known matched filter, and the threshold test of  $\tilde{\lambda}(y)$  against 0 leads to the optimal  $\mathcal{P}_e$  among all decision rules if  $\pi_0 = \pi_1$  [55]. The derivation of the  $\mathcal{R}$ -optimal Volterra filter here in fact resembles the historic derivation of the linear matched filter via maximizing an output signal-to-noise ratio [55]. The crucial differences are that here  $\tilde{\lambda}(y)$  can include higher-order products of  $y$  elements and the upper error bounds provide performance guarantees even for non-Gaussian statistics.

## B. Qubit readout

For an application of the detection theory, consider the qubit readout problem described in Refs. [20–22]. The goal is to infer the initial state of the qubit in one of the two possibilities from noisy measurements. The two hypotheses can be modeled as

$$\begin{aligned} \mathcal{H}_0 : y(t_k) &= y_0(t_k), \\ \mathcal{H}_1 : y(t_k) &= Sx(t_k) + y_0(t_k), \end{aligned} \quad (3.20)$$

where  $x$  is a hidden qubit observable that can undergo spontaneous decay or excitation in time,  $S$  is a positive signal amplitude, and  $y_0$  is a zero-mean noise process. To perform hypothesis testing given a record of  $y$ , consider the first-order  $\mathcal{R}$ -optimal decision rule given by

$$\tilde{H} = (\pi_0 C_0 + \pi_1 C_1)^{-1} \Delta, \quad (3.21)$$

$$\tilde{\lambda}(y) = \tilde{H}^\top (y - \bar{y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} 0, \quad (3.22)$$

where

$$\Delta = \frac{1}{2} (\langle y \rangle_1 - \langle y \rangle_0), \quad (3.23)$$

$$\bar{y} = \frac{1}{2} (\langle y \rangle_1 + \langle y \rangle_0), \quad (3.24)$$

$$C_0 = \langle yy^\top \rangle_0 - \langle y \rangle_0 \langle y \rangle_0^\top, \quad (3.25)$$

$$C_1 = \langle yy^\top \rangle_1 - \langle y \rangle_1 \langle y \rangle_1^\top, \quad (3.26)$$

and the upper error bounds are given by Eqs. (3.11)–(3.13).

$\tilde{\lambda}(y)$  for  $P = 1$  is a linear filter with respect to  $y$  and similar to the linear filters proposed in Ref. [20]. An advantage of the  $\mathcal{R}$ -optimal rule here is that the filter  $\tilde{H}$  depends only on the first-order moments  $\Delta(k)$  and  $\bar{y}(k)$  and second-order correlations  $C_0$  and  $C_1$ . All these moments can be simulated or measured directly in an experiment without the assumptions of continuous time, white Gaussian noise, and uncorrelated signal and noise made in prior work. The calculation of  $\tilde{H}$  is relatively straightforward compared with the numerical optimization procedure in Ref. [20], while  $\mathcal{Q}$  and  $\tilde{\mathcal{R}}$  provide theoretical performance guarantees. The upper bounds may be conservative, and a more precise comparison of  $\mathcal{P}_e(\tilde{H})$  with other linear or nonlinear filters [20–22] will require further numerical simulations and experimental tests.

To proceed further, consider the continuous-time limit. For the two-level  $x \in \{0, 1\}$  process with initial value  $x(0) = 1$  and spontaneous decay time  $T_1$  studied in Refs. [20, 22], it is not difficult [65] to show that the mean is

$$\langle x(t) \rangle = \exp\left(-\frac{t}{T_1}\right), \quad (3.27)$$

and the covariance function is

$$C_x(t, \tau) \equiv \langle x(t)x(\tau) \rangle - \langle x(t) \rangle \langle x(\tau) \rangle \quad (3.28)$$

$$= \exp\left[-\frac{\max(t, \tau)}{T_1}\right] - \exp\left(-\frac{t + \tau}{T_1}\right). \quad (3.29)$$

For a zero-mean white Gaussian noise with noise power  $\Pi$ ,

$$\langle y_0(t)y_0(\tau) \rangle = \Pi \delta(t - \tau). \quad (3.30)$$

The test statistic becomes

$$\tilde{\lambda} = \int_0^T dt \tilde{h}(t) \left[ y(t) - \frac{S}{2} \langle x(t) \rangle \right], \quad (3.31)$$

and a continuous-time limit of Eq. (3.21) leads to a Fredholm integral equation of the second kind [55] given by

$$\frac{S}{2} \langle x(t) \rangle = \Pi \tilde{h}(t) + \pi_1 S^2 \int_0^T d\tau C_x(t, \tau) \tilde{h}(\tau). \quad (3.32)$$

Further analytic simplifications may be possible for  $T \rightarrow \infty$  using Laplace transform, but a numerical solution of the Fredholm equation can easily be sought, as it is linear with respect to  $\tilde{h}_1$  and can be inverted in discrete time using, for example, the `mldivide` function in Matlab.

Define the input signal-to-noise ratio (SNR) as  $S^2 T_1 / \Pi$ . Fig. 1 plots some numerical examples of the filter for  $\pi_0 = \pi_1 = 1/2$  and  $T = 5T_1$ . The Matlab computation of all the filters shown with  $\delta t = 0.001T_1$  takes seconds to complete on a desktop PC. Fig. 2 plots the upper error bounds versus the input SNR. The upper bounds turn out to be conservative here, as a numerical investigation of  $\mathcal{P}_e$  later will demonstrate.

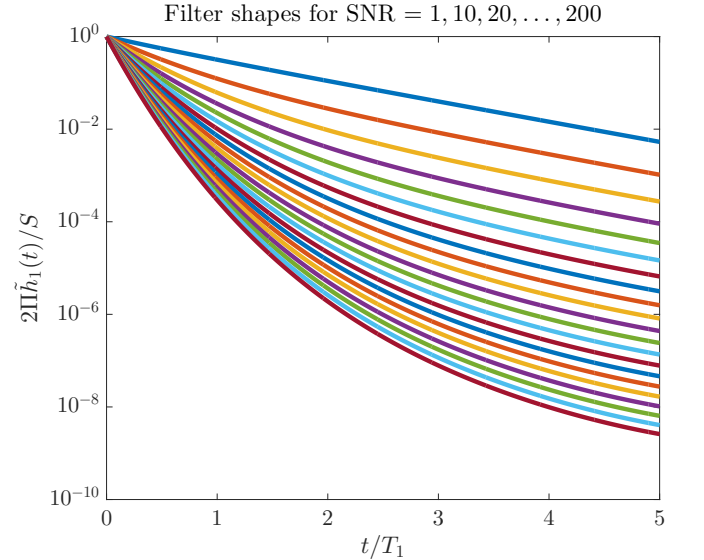


FIG. 1. (Color online). The normalized  $\mathcal{R}$ -optimal filters  $2\Pi\tilde{h}_1(t)/S$  in log scale versus normalized time  $t/T_1$  for different input SNR  $\equiv S^2 T_1 / \Pi = 1, 10, 20, \dots, 200$ .  $\pi_0 = \pi_1 = 1/2$  and  $T = 5T_1$  are assumed. The different plots can be distinguished by the reducing correlation times for increasing SNR.

The proposed decision rule can be compared with the optimal likelihood-ratio test (LRT) [55]. For the given problem, there exists an analytic expression for the log-



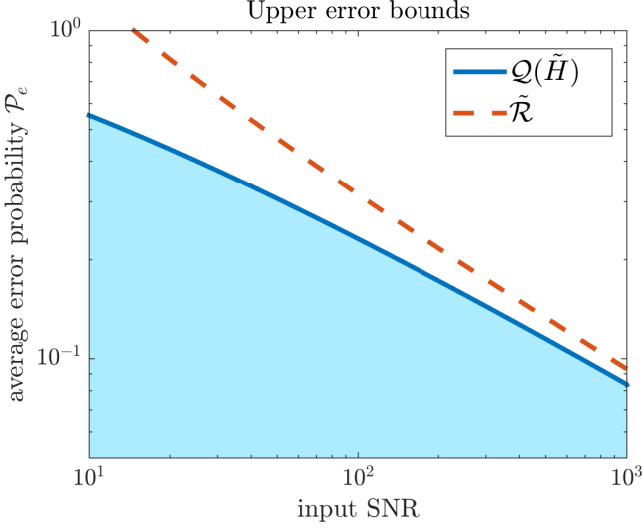


FIG. 2. (Color online). Upper bounds  $\mathcal{Q}(\tilde{H})$  and  $\tilde{\mathcal{R}}$  on the average error probability  $\mathcal{P}_e$  for the first-order Volterra filter versus input SNR from 10 dB to 30 dB in log-log scale.  $\pi_0 = \pi_1 = 1/2$  and  $T = 5T_1$  are assumed.  $\mathcal{P}_e$  is guaranteed to be in the shaded region below the curves.

likelihood ratio given by [22]

$$\lambda_o(y) = \frac{S}{\Pi} \int_0^T d\eta(t) \tilde{x}(t) - \frac{S^2}{2\Pi} \int_0^T dt \tilde{x}^2(t), \quad (3.33)$$

$$d\eta(t) = y(t)dt, \quad (3.34)$$

$$\tilde{x}(t) = \frac{p_1(t)}{p_0(t) + p_1(t)}, \quad (3.35)$$

$$p_1(t) = \exp \left[ \frac{S}{\Pi} \int_0^t d\eta(\tau) - t \left( \frac{S^2}{2\Pi} + \frac{1}{T_1} \right) \right], \quad (3.36)$$

$$p_0(t) = \frac{1}{T_1} \int_0^t d\tau p_1(\tau), \quad (3.37)$$

where the  $d\eta$  integrals are in the Itô sense. The optimal decision rule is thus

$$\lambda_o(y) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln \frac{\pi_0}{\pi_1}. \quad (3.38)$$

Although the LRT will achieve the lowest  $\mathcal{P}_e$ , the highly nonlinear dependence of  $\lambda_o$  on  $y$  makes its exact implementation difficult in real-time applications or for a large number of qubits.

The average error probabilities for both the  $\mathcal{R}$ -optimal rule and the LRT are estimated numerically using Monte Carlo simulations and plotted in Fig. 3. The errors are close at lower input SNR values. Considering the simplicity of the  $\mathcal{R}$ -optimal rule, the divergence at higher SNR is expected and indeed slight. At the input SNR of  $10^3$ ,  $\mathcal{P}_e$  for LRT is  $6.2 \times 10^{-3}$ , while that for the  $\mathcal{R}$ -optimal rule is only around a factor of 2 higher at  $1.48 \times 10^{-2}$ . A further optimization of  $\mathcal{P}_e$  beyond the results shown in Fig. 3 can be done by fine-tuning the threshold of the  $\mathcal{R}$ -optimal rule. For example, a numerical search for the

optimal threshold brings its error probability at input SNR =  $10^3$  down to  $8.3 \times 10^{-3}$ . A higher-order filter is hardly necessary for the SNRs considered here.

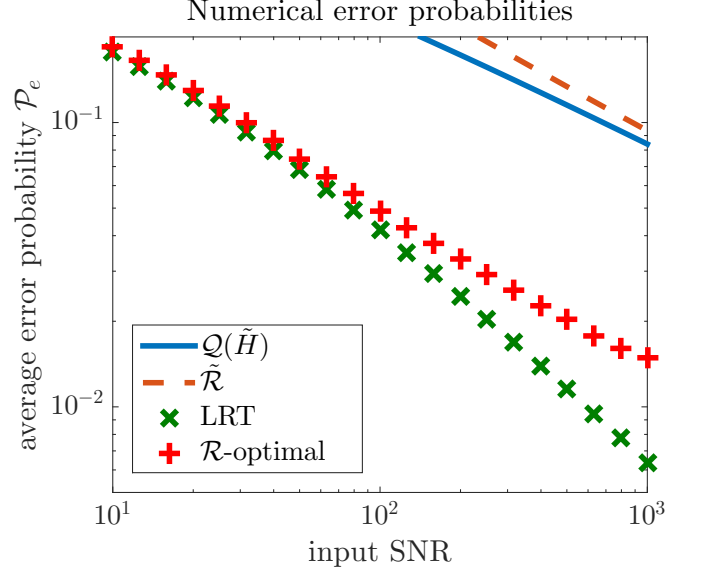


FIG. 3. (Color online). Numerically computed average error probabilities  $\mathcal{P}_e$  for the  $\mathcal{R}$ -optimal rule and the likelihood-ratio test (LRT) versus the input SNR from 10 dB to 30 dB in log-log scale.  $\pi_0 = \pi_1 = 1/2$  and  $T = 5T_1$  are assumed. Also shown are parts of the upper bounds  $\mathcal{Q}(\tilde{H})$  and  $\tilde{\mathcal{R}}$  for comparison.

The upper bounds depend only on low-order moments and apply equally to all problems with the same low-order moments, regardless of their higher-order statistics. It is not surprising that such indiscriminate bounds are loose for this particular example, as shown in Fig. 3. What is surprising is the near-optimal performance of a decision rule based on a loose upper bound. The log-likelihood ratio is given analytically for the problem considered here, so one may compare it with the  $\mathcal{R}$ -optimal test statistic to see how the two resemble each other. In general, however, the log-likelihood ratio is difficult or even impossible to compute if the full probability models are more complicated or simply unidentified. The  $\mathcal{R}$ -optimal rule requires only low-order moments to be known, and is hence more convenient to implement in practice.

#### IV. CONCLUSION

I have proposed the use of Volterra filters for quantum estimation and detection. The importance of the proposal lies in its promise to solve many of the practical problems associated with existing optimal quantum inference techniques, including the curse of dimensionality, the lack of performance assurances upon approximations, and the need for a Markovian model. Beyond the

examples of quantum state tomography and qubit readout discussed in this paper, diverse applications in quantum information processing [1, 38, 51], including cooling [26], squeezing [27], state preparation [28], metrology [6–8, 16–18, 42–44], fundamental tests of quantum mechanics [18, 19, 38, 61], and error correction [29, 30], are expected to benefit. Potential extensions of the theory include adaptive, recursive, and coherent generalizations for feedback control [1] and noise cancellation [66], filter training via machine learning [67], robustness analysis, the use of other performance criteria for improved robustness [68] or multi-hypothesis testing [18, 19], a connection with Shannon information theory through the

relations between filtering errors and entropic information [69], and a study of fundamental uncertainty relations in conjunction with quantum lower error bounds [6, 7, 16, 42–44].

## ACKNOWLEDGMENTS

This work is supported by the Singapore National Research Foundation under NRF Grant No. NRF-NRFF2011-07.

- 
- [1] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).
  - [2] C. W. Gardiner and P. Zoller, *Quantum Noise* (Springer-Verlag, Berlin, 2004).
  - [3] K. Jacobs, *Quantum Measurement Theory and its Applications* (Cambridge University Press, Cambridge, 2014).
  - [4] L. Bouten, R. Van Handel, and M. James, *SIAM Journal on Control and Optimization* **46**, 2199 (2007); L. Bouten, R. van Handel, and M. R. James, *SIAM Review* **51**, 239 (2009).
  - [5] S. Haroche and J. M. Raimond, *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford University Press, Oxford, 2006).
  - [6] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
  - [7] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer-Verlag, Berlin, 2001).
  - [8] V. B. Braginsky and F. Y. Khalili, *Quantum Measurement* (Cambridge University Press, Cambridge, 1992).
  - [9] M. G. A. Paris and J. Řeháček, eds., *Quantum State Estimation* (Springer-Verlag, Berlin, 2004).
  - [10] R. Blume-Kohout, *New Journal of Physics* **12**, 043034 (2010); C. Ferrie, *New Journal of Physics* **16**, 093035 (2014).
  - [11] C. A. Riofrío, *Continuous Measurement Quantum State Tomography of Atomic Ensembles*, Ph.D. thesis, University of New Mexico, Albuquerque (2014).
  - [12] R. L. Cook, C. A. Riofrío, and I. H. Deutsch, *Phys. Rev. A* **90**, 032113 (2014).
  - [13] K. M. R. Audenaert and S. Scheel, *New Journal of Physics* **11**, 023028 (2009).
  - [14] P. Six, P. Campagne-Ibarcq, I. Dotsenko, A. Sarlette, B. Huard, and P. Rouchon, ArXiv e-prints (2015), [arXiv:1510.01726 \[quant-ph\]](#).
  - [15] C. Granade, J. Combes, and D. G. Cory, ArXiv e-prints (2015), [arXiv:1509.03770 \[quant-ph\]](#).
  - [16] V. Giovannetti, S. Lloyd, and L. Maccone, *Nature Photon.* **5**, 222 (2011).
  - [17] M. Tsang, *Phys. Rev. Lett.* **102**, 250403 (2009); *Phys. Rev. A* **80**, 033840 (2009); *Phys. Rev. A* **81**, 013824 (2010); S. Gammelmark, B. Julsgaard, and K. Mølmer, *Phys. Rev. Lett.* **111**, 160401 (2013); I. Guevara and H. Wiseman, ArXiv e-prints (2015), [arXiv:1503.02799 \[quant-ph\]](#).
  - [18] M. Tsang, *Phys. Rev. Lett.* **108**, 170502 (2012).
  - [19] M. Tsang, *Quantum Meas. Quantum Metr.* **1**, 84 (2013).
  - [20] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* **76**, 012325 (2007).
  - [21] B. D’Anjou and W. A. Coish, *Phys. Rev. A* **89**, 012313 (2014); B. D’Anjou, L. Kurek, L. Childress, and W. A. Coish, ArXiv e-prints (2015), [arXiv:1507.06846 \[quant-ph\]](#).
  - [22] S. Ng and M. Tsang, *Phys. Rev. A* **90**, 022325 (2014).
  - [23] T. A. Wheatley, D. W. Berry, H. Yonezawa, D. Nakane, H. Arao, D. T. Pope, T. C. Ralph, H. M. Wiseman, A. Furusawa, and E. H. Huntington, *Phys. Rev. Lett.* **104**, 093601 (2010); H. Yonezawa, D. Nakane, T. A. Wheatley, K. Iwasawa, S. Takeda, H. Arao, K. Ohki, K. Tsumura, D. W. Berry, T. C. Ralph, H. M. Wiseman, E. H. Huntington, and A. Furusawa, *Science* **337**, 1514 (2012); K. Iwasawa, K. Makino, H. Yonezawa, M. Tsang, A. Davidovic, E. Huntington, and A. Furusawa, *Phys. Rev. Lett.* **111**, 163602 (2013).
  - [24] V. P. Belavkin, *Physics Letters A* **140**, 355 (1989).
  - [25] V. P. Belavkin, ArXiv Mathematical Physics e-prints (2007), [arXiv:math-ph/0702079](#), and references therein.
  - [26] D. A. Steck, K. Jacobs, H. Mabuchi, T. Bhattacharya, and S. Habib, *Phys. Rev. Lett.* **92**, 223004 (2004); D. A. Steck, K. Jacobs, H. Mabuchi, S. Habib, and T. Bhattacharya, *Phys. Rev. A* **74**, 012322 (2006).
  - [27] L. K. Thomsen, S. Mancini, and H. M. Wiseman, *Phys. Rev. A* **65**, 061801 (2002).
  - [28] M. Yanagisawa, *Phys. Rev. Lett.* **97**, 190201 (2006); A. Negretti, U. V. Poulsen, and K. Mølmer, *Phys. Rev. Lett.* **99**, 223601 (2007).
  - [29] C. Ahn, A. C. Doherty, and A. J. Landahl, *Phys. Rev. A* **65**, 042301 (2002); M. Sarovar, C. Ahn, K. Jacobs, and G. J. Milburn, *Phys. Rev. A* **69**, 052324 (2004).
  - [30] B. A. Chase, A. J. Landahl, and J. M. Geremia, *Phys. Rev. A* **77**, 032304 (2008).
  - [31] D. B. Hume, T. Rosenband, and D. J. Wineland, *Phys. Rev. Lett.* **99**, 120502 (2007).
  - [32] C. Sayrin, I. Dotsenko, X. Zhou, B. Peaudecerf, T. Rybarczyk, S. Gleyzes, P. Rouchon, M. Mirrahimi, H. Amini, M. Brune, J.-M. Raimond, and S. Haroche, *Nature (London)* **477**, 73 (2011).
  - [33] W. Wieczorek, S. G. Hofer, J. Hoelscher-Obermaier, R. Riedinger, K. Hammerer, and M. Aspelmeyer, *Phys. Rev. Lett.* **114**, 223601 (2015).

- [34] I. G. Vladimirov and I. R. Petersen, ArXiv e-prints (2012), [arXiv:1202.0946 \[quant-ph\]](#).
- [35] H. Amini, R. A. Somaraju, I. Dotsenko, C. Sayrin, M. Mirrahimi, and P. Rouchon, *Automatica* **49**, 2683 (2013).
- [36] M. R. Hush, S. S. Szigeti, A. R. R. Carvalho, and J. J. Hope, *New Journal of Physics* **15**, 113060 (2013).
- [37] A. E. B. Nielsen, A. S. Hopkins, and H. Mabuchi, *New Journal of Physics* **11**, 105043 (2009).
- [38] M. Aspelmeyer, T. J. Kippenberg, and F. Marquardt, *Rev. Mod. Phys.* **86**, 1391 (2014).
- [39] I. Bloch, J. Dalibard, and W. Zwerger, *Rev. Mod. Phys.* **80**, 885 (2008).
- [40] A. A. Houck, H. E. Tureci, and J. Koch, *Nature Phys.* **8**, 292 (2012).
- [41] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing* (Wiley, New York, 2000).
- [42] M. Tsang, H. M. Wiseman, and C. M. Caves, *Phys. Rev. Lett.* **106**, 090401 (2011).
- [43] M. Tsang, *Phys. Rev. Lett.* **108**, 230401 (2012); D. W. Berry, M. Tsang, M. J. W. Hall, and H. M. Wiseman, *Phys. Rev. X* **5**, 031018 (2015).
- [44] M. Tsang and R. Nair, *Phys. Rev. A* **86**, 042115 (2012); M. Tsang, *New Journal of Physics* **15**, 073005 (2013).
- [45] J. Zhang, Y.-X. Liu, R.-B. Wu, K. Jacobs, S. Kaya Ozdemir, L. Yang, T.-J. Tarn, and F. Nori, ArXiv e-prints (2014), [arXiv:1407.8108 \[quant-ph\]](#).
- [46] V. P. Belavkin, *Foundations of Physics* **24**, 685 (1994).
- [47] Q. Gao, D. Dong, and I. R. Petersen, ArXiv e-prints (2015), [arXiv:1504.06780 \[math-ph\]](#).
- [48] M. Ozawa, *Phys. Rev. A* **67**, 042105 (2003).
- [49] J. S. Bell, *Speakable and Unsayable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987); R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [50] A. J. Leggett and A. Garg, *Phys. Rev. Lett.* **54**, 857 (1985); C. Emary, N. Lambert, and F. Nori, *Reports on Progress in Physics* **77**, 016001 (2014).
- [51] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2011).
- [52] U. Weiss, *Quantum Dissipative Systems* (World Scientific, Singapore, 2008); S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995); H. Bruus and K. Flensberg, *Introduction to Many-Body Quantum Theory in Condensed Matter Physics* (Oxford University Press, Oxford, 2004).
- [53] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).
- [54] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).
- [55] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. (John Wiley & Sons, New York, 2001).
- [56] J. Stockton, M. Armen, and H. Mabuchi, *J. Opt. Soc. Am. B* **19**, 3019 (2002).
- [57] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (Wiley, Hoboken, 2006).
- [58] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, *Rev. Mod. Phys.* **82**, 1155 (2010).
- [59] H. Miao, Y. Ma, C. Zhao, and Y. Chen, ArXiv e-prints (2015), [arXiv:1506.00117 \[quant-ph\]](#).
- [60] S. T. Flammia and Y.-K. Liu, *Phys. Rev. Lett.* **106**, 230501 (2011).
- [61] Y. Chen, *Journal of Physics B: Atomic, Molecular and Optical Physics* **46**, 104001 (2013).
- [62] B. Picinbono and P. Duvaut, *IEEE Transactions on Information Theory* **36**, 1061 (1990).
- [63] B. Picinbono, *IEEE Transactions on Aerospace and Electronic Systems* **31**, 1072 (1995).
- [64] P. Billingsley, *Probability and Measure* (Wiley, New York, 1995).
- [65] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, Berlin, 2010).
- [66] M. Tsang and C. M. Caves, *Phys. Rev. Lett.* **105**, 123601 (2010); *Phys. Rev. X* **2**, 031016 (2012); N. Yamamoto, *Phys. Rev. X* **4**, 041029 (2014).
- [67] A. Hentschel and B. C. Sanders, *Phys. Rev. Lett.* **104**, 063603 (2010); E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, *Phys. Rev. Lett.* **114**, 200501 (2015).
- [68] M. R. James, *Phys. Rev. A* **69**, 032108 (2004).
- [69] A. Barchielli and G. Lupieri, “Information gain in quantum continual measurements,” in *Quantum Stochastics and Information*, edited by V. P. Belavkin and M. Gu (World Scientific, Singapore, 2008) Chap. 15, pp. 325–345, [arXiv:quant-ph/0612010](#); M. Tsang, in *2014 IEEE International Symposium on Information Theory (ISIT)* (2014) pp. 321–325.